CEBS: Chemical Effects in Biological Systems.
An integrated data management system for the NTP

Jennifer M. Fostel, Ph.D.
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors

June 21-22, 2010

CEBS = Chemical Effects in Biological Systems

- The CEBS database includes more than responses to chemicals:
  - Studies of ***chemical*** test articles
  - Studies of ***environmental agents*** such as ozone, hyperoxia
  - Studies of the responses of ***genetic changes*** such as knockouts
  - Studies of effects of ***physical agents*** such as magnetic fields

- CEBS was originally developed by NIEHS Division of Intramural Research to house data of interest to toxicologists and environmental health scientists.
  - DIR scientists conducting toxicogenomics and proteomics studies
  - Public microarray datasets developed by industry and academic labs
  - Result: CEBS has a flexible design, open to a variety of study types

CEBS – from DIR to NTP

- Because CEBS can house data from many different study designs, it is well-suited to house *all* NTP data in a single database.
- Once data are in CEBS they are integrated. Thus the data can be queried on a per-study OR per-compound basis.
- CEBS will be used to:
  - Perform cross-study searches and analysis on NTP data
  - Serve (public) NTP data to the public
  - Permit NTP data to be integrated with other reference datasets

Content: Comparison of CEBS and other databases

	Protocol details	Microarray data	Clinical pathology histopathology data	Reproductive toxicology data	Conclusions	Integrated data
Reference toxicology databases				√		
Microarray repositories	√	√			√	
NTP databases	√		√	√	√	
CEBS	√ √	√	√	√	√	√

Aim: To house all public NTP data in CEBS

- Components of this task:
  - Collect and load NTP legacy data into CEBS
  - Set up processes to load data from on-going studies
  - Modify CEBS user interface to highlight features of NTP studies

CEBS content – current status

- Load NTP microarray data (completed; 3 studies published)
- Load legacy data from NTP databases
  - Genetic toxicology results (in progress)
  - Clinical pathology data and immunotoxicology data (completed)
  - Developmental and reproductive toxicology (in progress)
  - Bioassay data (next up)
- Load data from on-going studies:
  - Collect high-throughput screening data from NIH Chemical Genomics Center (completed to date)
  - Align with Project Officers to collect interim and final data from labs (process in place)

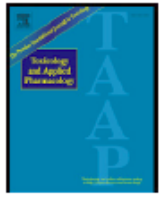
Example of using CEBS –
Show me the data from an NTP study

Toxicology and Applied Pharmacology 243 (2010) 300–314

Contents lists available at ScienceDirect

Toxicology and Applied Pharmacology

journal homepage: www.elsevier.com/locate/ytaap



Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning

Scott S. Auerbach [a], Ruchir R. Shah [b], Deepak Mav [b], Cynthia S. Smith [a], Nigel J. Walker [a], Molly K. Vallant [a], Gary A. Boorman [a], Richard D. Irwin [a,*]

[a] National Toxicology Program, National Institute of Environmental Health Sciences, NIH, RTP, NC 27709, USA
[b] SRA International, RTP, NC 27709, USA

ARTICLE INFO

Article history:
Received 4 September 2009
Revised 18 November 2009
Accepted 20 November 2009
Available online 11 December 2009

Keywords:
Toxicogenomics
Cancer
Liver
Prediction
Alkenylbenzene
Rat

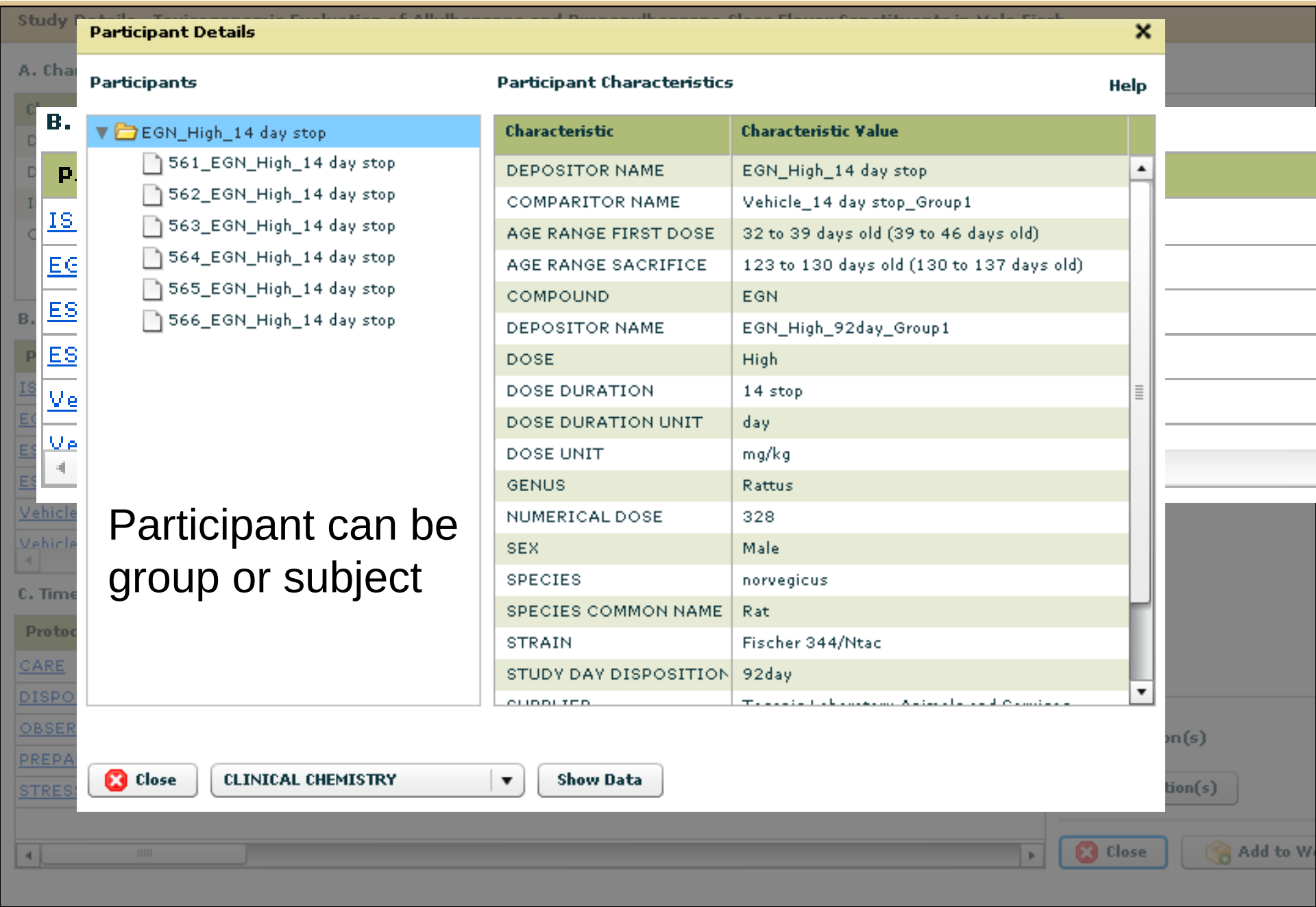
ABSTRACT

Identification of carcinogenic activity is the primary goal of the 2-year bioassay. The expense of these studies limits the number of chemicals that can be studied and therefore chemicals need to be prioritized based on a variety of parameters. We have developed an ensemble of support vector machine classification models based on male F344 rat liver gene expression following 2, 14 or 90 days of exposure to a collection of hepatocarcinogens (aflatoxin B1, 1-amino-2,4-dibromoanthraquinone, *N*-nitrosodimethylamine, methyleugenol) and non-hepatocarcinogens (acetaminophen, ascorbic acid, tryptophan). Seven models were generated based on individual exposure durations (2, 14 or 90 days) or a combination of exposures (2 + 14, 2 + 90, 14 + 90 and 2 + 14 + 90 days). All sets of data, with the exception of one yielded models with 0% cross-validation error. Independent validation of the models was performed using expression data from the liver of rats exposed at 2 dose levels to a collection of alkenylbenzene flavoring agents. Depending on the model used and the exposure duration of the test data, independent validation error rates ranged from 47% to 10%. The variable with the most notable effect on independent validation accuracy was exposure duration of the alkenylbenzene test data. All models generally exhibited improved performance as the exposure duration of the

All Data

Show Details | Add to Workspace

Investigation/Study	Accession Number
Characterization of NCI60 cell lines.	007-00001-0010-00
Characterization of recombinant in-bred mouse strains	003-00001-0010-00
Dose-and time-responses to acute administration of acetaminophen	002-00001-0010-00
Effects of Acetaminophen - NCT Microarray Investigation	001-00002-0010-00
Effects of aryl hydrocarbon receptor (AhR) ligands on hepatic gene expression.	003-00004-0001-00
Effects of phenobarbital	004-00001-0010-00
Expression of marker genes in mouse L5178Y cells or human TK6 cells following chemical exposure.	008-00004-0001-00
Gene expression patterns in response to a panel of hepatocarcinogens	004-00005-0010-00
Gene expression profiles in the cerebellum and hippocampus following exposure to a neurotoxicant, Aroclor 1254	010-00001-0001-00
Genomic biomarkers to predict increased lung tumor incidence in 2-year rodent cancer bioassays	004-00006-0010-00
HESI Baseline Animal Data Library	008-00003-0001-00
HESI Hepatotoxicity Investigation	008-00001-0010-00
HESI Nephrotoxicity Investigation	008-00002-0010-00
Prediction of hepatocarcinogenic potential of alkylbenzene flavoring agents using transcriptomics and machin	
Toxicogenomic Evaluation of Allylbenzene and Propenylbenzene Class Flavor Constituents in Male Fischer	
Toxicogenomic Evaluation of Rat Liver Carcinogens and Non-carcinogens in Male Fischer 344 Rats	
Pathways regulated by toll-like receptor 4 (TLR4)	005-00003-0030-00
Prediction of hepatocarcinogenic potential of alkylbenzene flavoring agents using transcriptomics and machine learning.	002-00100-0001-00
Protective Role of IL-10 in Ozone-Induced Pulmonary Inflammation	005-00003-0070-00
Selection for drug-resistance increases the number of cells with cancer stem cell characteristics	007-00002-0010-00
TRC Acetaminophen Standardization	009-00001-0010-00
Tissue Atlas	004-00007-0000-00
US EPA, ORD, Small Fish Computational Toxicology - Zebrafish (Phase 2) Exposures to Endocrine Active Compounds with Differing Modes of Action	010-00002-0001-00


Participant Details
Participants
Participant Characteristics
Help
EGN_High_14 day stop
561_EGN_High_14 day stop
562_EGN_High_14 day stop
563_EGN_High_14 day stop
564_EGN_High_14 day stop
565_EGN_High_14 day stop
566_EGN_High_14 day stop
Characteristic
Characteristic Value
DEPOSITOR NAME
EGN_High_14 day stop
COMPARITOR NAME
Vehicle_14 day stop_Group1
AGE RANGE FIRST DOSE
32 to 39 days old (39 to 46 days old)
AGE RANGE SACRIFICE
123 to 130 days old (130 to 137 days old)
COMPOUND
EGN
DEPOSITOR NAME
EGN_High_92day_Group1
DOSE
High
DOSE DURATION
14 stop
DOSE DURATION UNIT
day
DOSE UNIT
mg/kg
GENUS
Rattus
NUMERICAL DOSE
328
SEX
Male
SPECIES
norvegicus
SPECIES COMMON NAME
Rat
STRAIN
Fischer 344/Ntac
STUDY DAY DISPOSITION
92day
Participant can be group or subject
Close
CLINICAL CHEMISTRY
Show Data

Study Protocol Details (CARE)

Study Protocol Details (STRESSOR_PROTOCOL)

Protocol Type/Name/Attributes	Value
ANT_high	
VOLUME PER ADMIN UNIT	mL/kg
BIOLOGICAL DOSE RANGE	High
BIOLOGICAL CLASS AGENT	Non-carcinogen
TIME OF DAY OF DOSE	8 AM and noon; 8 AM and 11 AM on da
PURITY MEASUREMENT METHOD	gas chromatography with flame ionizatio
PURITY PERCENTAGE	98.6
STORAGE TEMP UNIT	degree C
STORAGE TEMP	20
CHEM SUPPLIER	MRI
STRESSOR NAME	Anethole
VOLUME PER ADMIN	5
DOSE FREQUENCY DESC	daily on weekdays; ensure 5 consercuti
DOSE FREQUENCY UNIT	per day

Close

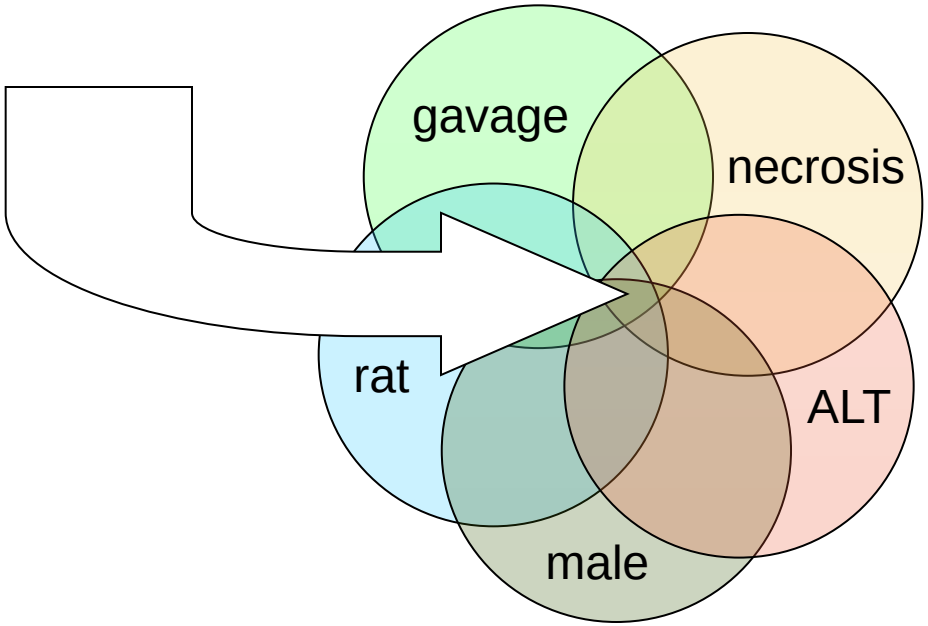
CLINICAL CHEMISTRY

BILE (UM/L)	BILE (UM/L)	BUN (MG/DL)	BUN (MG/DL)	CHOL (MG/DL)	CHOL (MG/DL)	CK (U/L)	CK (U/L)	CREA (MG/
22	22	12	12	108	108	278	278	0.4
28	28	12	12	96	96	382	382	0.4
8	8	12	12	84	84	181	181	0.6
10	10	16	16	102	102	320	320	0.6
15	15	17	17	103	103	225	225	0.6
3	3	14	14	94	94	264	264	0.6
6	6	12	12	118	118	3851	3851	0.3
15	15	11	11	73	73	303	303	0.6
6	6	14	14	84	84	233	233	0.6
33	33	13	13	96	96	393	393	0.5
17	17	11	11	84	84	310	310	0.5
6	6	14	14	80	80	308	308	0.6
19	19	11	11	74	74	103	103	0.6
11	11	15	15	81	81	78	78	0.5
13	13	13	13	83	83	247	247	0.5

Close Download File Size: 23 KB

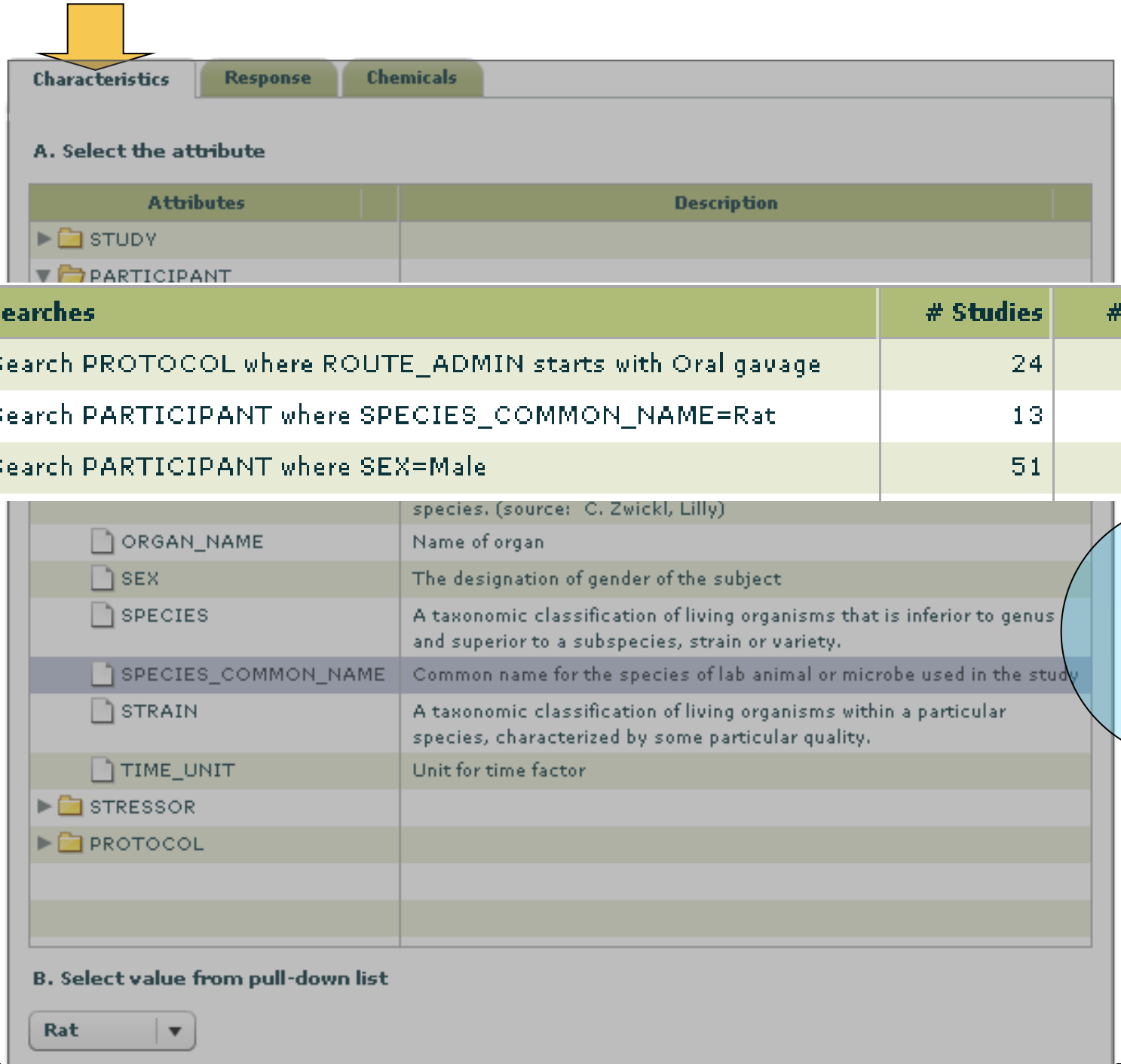
CEBS Example #2 – Identify genes with altered expression in livers from rats experiencing hepatotoxicity

- Part 1 – select rats of interest
  - Select males, rats, and animals treated via gavage
  - Select animals showing elevated alanine aminotransferase (ALT) and animals with marked centrilobular necrosis
  - Take intersection:

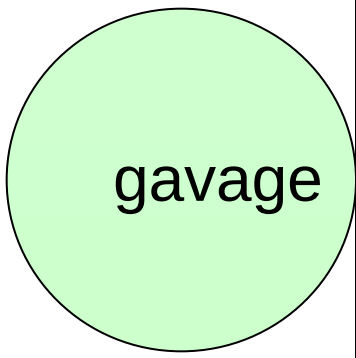


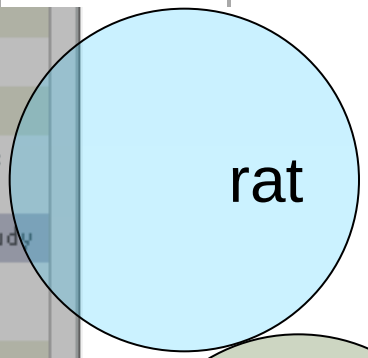
CEBS Example #2 – Identify genes with altered expression in livers from rats experiencing hepatotoxicity

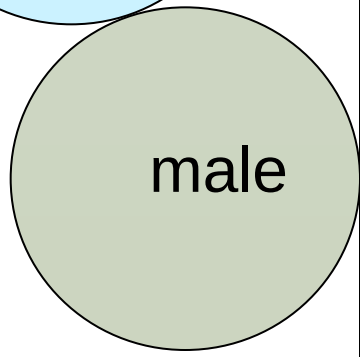
- Remaining steps in example –
  - Find comparator animals using CEBS
  - Retrieve microarray data from selected animals and comparators
  - Carry out t-test (arrays from comparator animals vs selected)
  - Export results

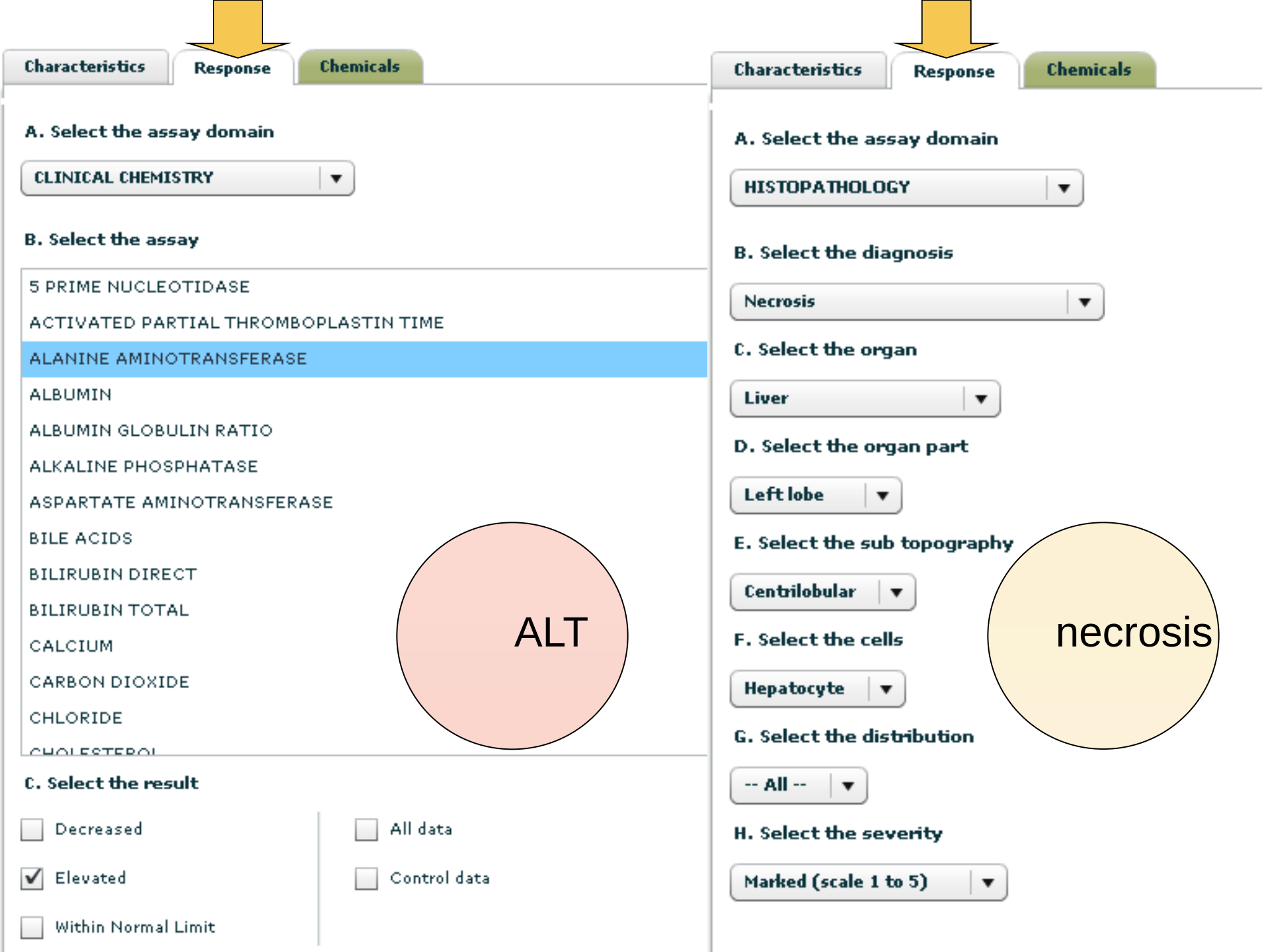


Searches	# Studies	# Groups	# Subjects
Search PROTOCOL where ROUTE_ADMIN starts with Oral gavage	24	346	1364
Search PARTICIPANT where SPECIES_COMMON_NAME=Rat	13	339	1986
Search PARTICIPANT where SEX=Male	51	2510	8191



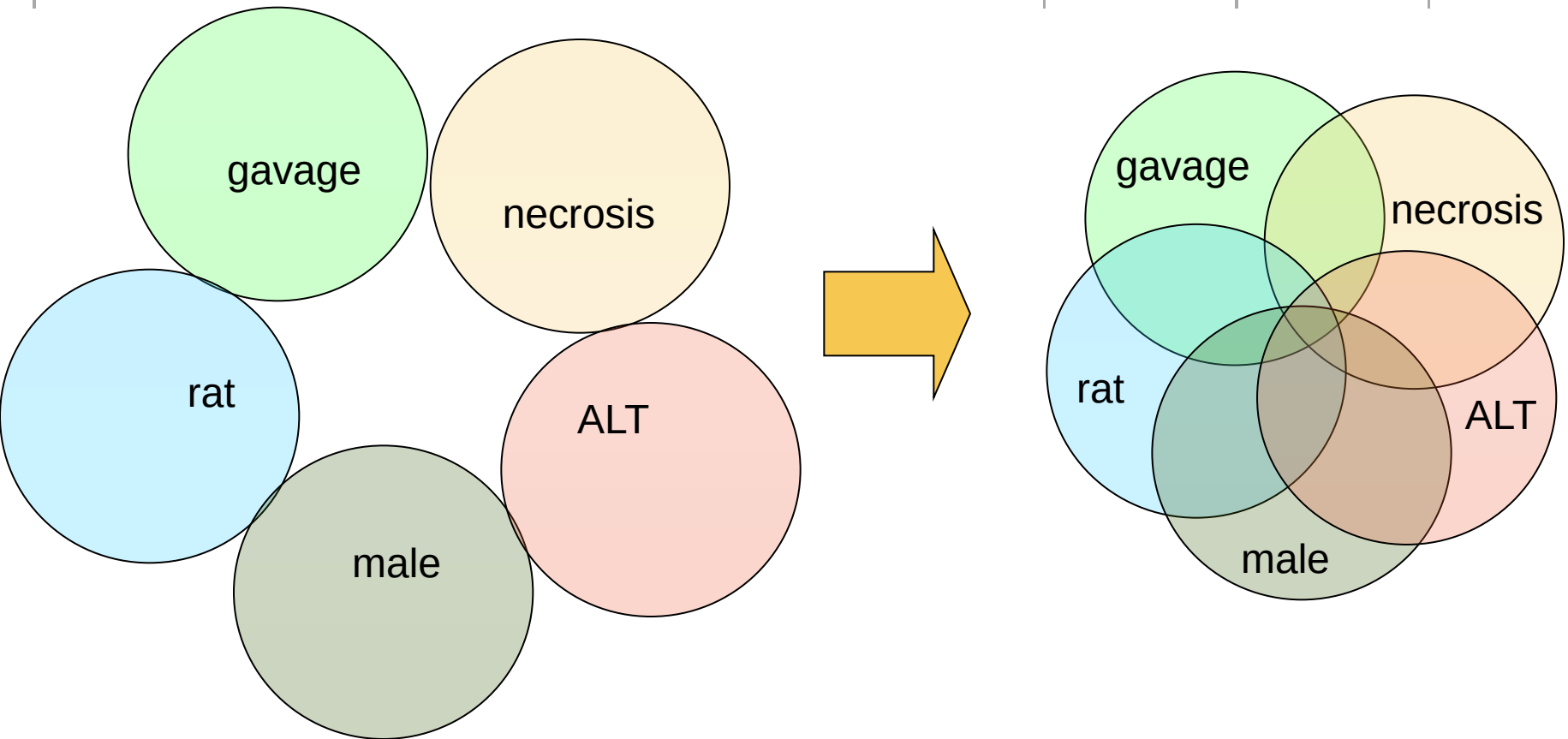





Characteristics
Response
Chemicals
A. Select the assay domain
CLINICAL CHEMISTRY
B. Select the assay
5 PRIME NUCLEOTIDASE
ACTIVATED PARTIAL THROMBOPLASTIN TIME
ALANINE AMINOTRANSFERASE
ALBUMIN
ALBUMIN GLOBULIN RATIO
ALKALINE PHOSPHATASE
ASPARTATE AMINOTRANSFERASE
BILE ACIDS
BILIRUBIN DIRECT
BILIRUBIN TOTAL
CALCIUM
CARBON DIOXIDE
CHLORIDE
ALT
C. Select the result
Decreased
Elevated
Within Normal Limit
All data
Control data
Characteristics
Response
Chemicals
A. Select the assay domain
HISTOPATHOLOGY
B. Select the diagnosis
Necrosis
C. Select the organ
Liver
D. Select the organ part
Left lobe
E. Select the sub topography
Centrilobular
F. Select the cells
Hepatocyte
G. Select the distribution
-- All --
H. Select the severity
Marked (scale 1 to 5)
necrosis

Results of the five searches:

Searches	# Studies	# Groups	# Subjects
Search PROTOCOL where ROUTE_ADMIN starts with Oral gavage	24	346	1364
Search PARTICIPANT where SPECIES_COMMON_NAME=Rat	13	339	1986
Search PARTICIPANT where SEX=Male	51	2510	8191
Search RESPONSE where ALANINE_AMINOTRANSFERASE=ELEVATI	22	83	236
Search RESPONSE where HISTOPATHOLOGY observerations have Necr	7	15	44



Combine searches in CEBS

A. Enter name for combined search

Male rats, gavaged, with inc. ALT and necrosis

B. Select two or more searches to combine

(Hold shift key for multi-select)

Search PROTOCOL where ROUTE_ADMIN starts with gavage

Search PARTICIPANT where SPECIES_COMMON_NAME=Rat

Search PARTICIPANT where SEX=Male

Search RESPONSE where ALANINE_AMINOTRANSFERASE=ELEVATED|

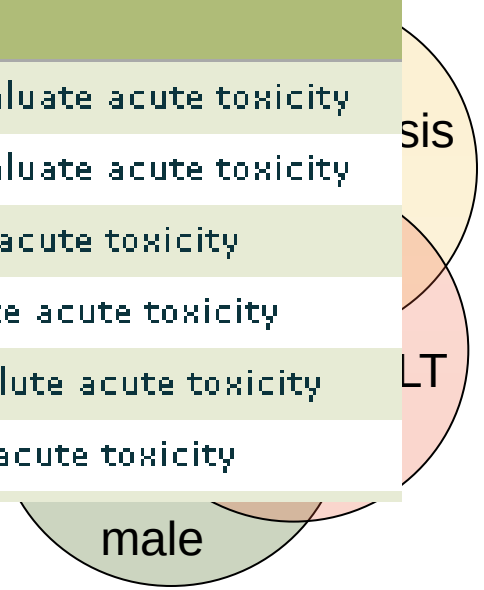
Search RESPONSE where HISTOPATHOLOGY obsverations have Necrosis|L

C. Select

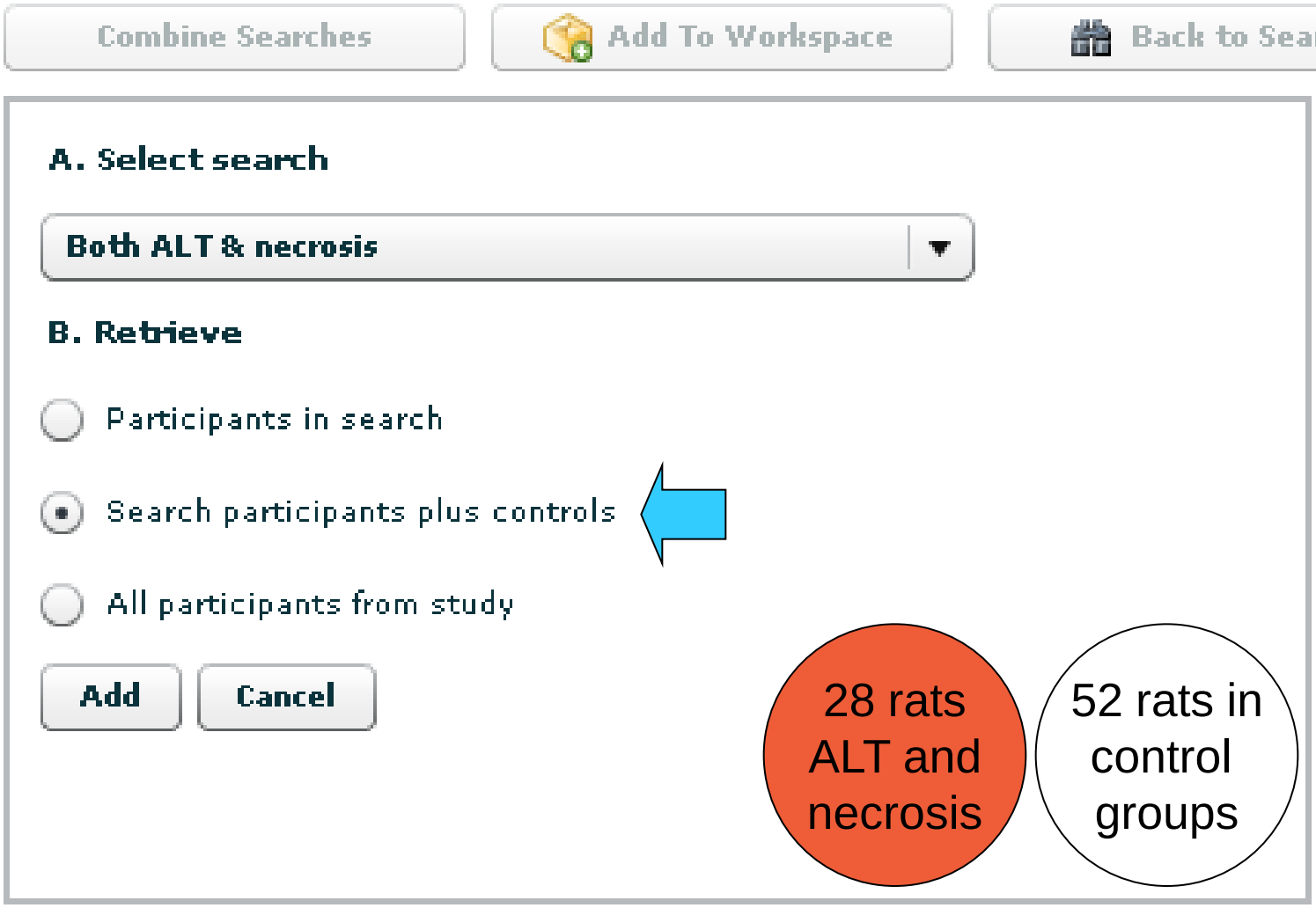
Cancel

Searches	# Studies	# Groups	# Subjects
Search PROTOCOL where ROUTE_ADMIN starts	24	346	1364
Search PARTICIPANT where SPECIES_COMMON	13	339	1986
Search PARTICIPANT where SEX=Male	51	2510	8191
Search RESPONSE where ALANINE_AMINOTRAN	22	83	236
Search RESPONSE where HISTOPATHOLOGY ob:	7	15	44
Male rats, gavaged, with inc. ALT and necrosis	6	10	28

Study Title
Application of 1,2-dichlorobenzene to F344 rats via oral gavage to evaluate acute toxicity
Application of 1,4-dichlorobenzene to F344 rats via oral gavage to evaluate acute toxicity
Application of bromobenzene to F344 rats via oral gavage to evaluate acute toxicity
Application of monocrotaline to F344/N rats via oral gavage to evaluate acute toxicity
Application of n-nitrosomorpholine to F344 rats via oral gavage to evalute acute toxicity
Application of thioacetamide to F344 rats via oral gavage to evaluate acute toxicity



Selected rats plus comparators (identified by depositor)



Use CEE

Grou

Export re

List of ge
significar
change r
hepatoto

CEBS_PROBE_ID	GENE_NAME	GROUP_A	GROUP_E	TWO_SIDE	LogP
1390672_at	reprimo, TP53 dependent G2 arrest mediator candidate	3.992	6.126	9.06E-21	-20.043
1375186_at	DPH3, KTI11 homolog (S. cerevisiae)	9.96	8.63	3.76E-20	-19.4252
1374883_at	myotubularin related protein 7	4.47	6.637	5.49E-20	-19.2608
1370244_at	cathepsin L1	13.158	12.086	1.57E-19	-18.8042
1390097_at	TSPY-like 4	5.271	6.967	2.87E-19	-18.5425
1371486_at	similar to U1 small nuclear ribonucleoprotein C (U1 snR	10.167	9.561	5.69E-19	-18.2446
1371486_at	small nuclear ribonucleoprotein polypeptide C	10.167	9.561	5.69E-19	-18.2446
1374135_at	importin 4	9.117	7.979	1.08E-18	-17.9683
1370583_s_at	ATP-binding cassette, sub-family B (MDR	8.629	5.248	2.43E-18	-17.6143
1370583_s_at	TAP), member 1A	8.629	5.248	2.43E-18	-17.6143
1370583_s_at	TAP), member 1B	8.629	5.248	2.43E-18	-17.6143
1371777_at	poly A binding protein, cytoplasmic 4	10.155	8.605	4.97E-18	-17.3033
1388516_at	LRRGT00141	10.394	9.445	5.37E-18	-17.2698
1370355_at	stearoyl-Coenzyme A desaturase 1	6.344	12.074	7.38E-18	-17.132
1388378_at	eukaryotic translation initiation factor 3, subunit C	10.864	9.799	7.95E-18	-17.0996
1398839_at	thioredoxin 1	13.229	12.582	1.40E-17	-16.855
1370838_s_at	alpha-spectrin 2	9.906	8.762	1.69E-17	-16.7732
1367565_a_at	ferritin, heavy polypeptide 1	14.055	13.451	1.72E-17	-16.7643
1375522_at	N-myristoyltransferase 1	9.497	8.673	2.02E-17	-16.6955
1372124_at	eukaryotic translation initiation factor 4B	8.37	7.455	2.26E-17	-16.6455
1372092_at	trafficking protein, kinesin binding 2	7.171	6.347	3.27E-17	-16.4857
1371237_a_at	metallothionein 1a	12.497	8.261	3.30E-17	-16.4814
1373955_at	importin 5	9.474	7.563	5.85E-17	-16.2326
1368486_at	insulin receptor substrate 3	4.288	6.001	6.33E-17	-16.1984
1386926_at	acyl-CoA synthetase long-chain family member 5	10.408	11.351	8.25E-17	-16.0835
1367559_at	ferritin, light polypeptide	13.197	12.418	8.39E-17	-16.076
1371539_at	nucleolar protein family A, member 2	9.908	8.28	9.05E-17	-16.0435
1371384_at	basic transcription factor 3	11.597	10.759	1.08E-16	-15.9683
1369930_at	proteasome (prosome, macropain) subunit, alpha type 6	12.338	11.656	1.42E-16	-15.8484
1383625_a_at	zinc finger protein 259	9.327	7.912	1.49E-16	-15.8259
1386866_at	tryptophan 5-monooxygenase activation protein, gamma	10.461	9.214	1.58E-16	-15.8
1386866_at	tyrosine 3-monooxygenase	10.461	9.214	1.58E-16	-15.8
1388271_at	metallothionein 2A	11.638	7.889	1.75E-16	-15.7558
1388390_at	eukaryotic translation initiation factor 3, subunit H	10.757	10.155	1.87E-16	-15.7292
1388110_at	eukaryotic translation elongation factor 1 alpha 1	13.528	12.999	2.04E-16	-15.6911
1388110_at	similar to eukaryotic translation elongation factor 1 alph	13.528	12.999	2.04E-16	-15.6911
1371641_at	chaperonin containing Tcp1, subunit 7 (eta)	10.899	9.748	2.66E-16	-15.5746
1388576_at	eukaryotic translation initiation factor 3, subunit 9 (eta)	9.563	8.486	3.56E-16	-15.4486
1370277_at	solute carrier family 25 (mitochondrial carrier, phosphat	12.235	11.672	3.61E-16	-15.4424

Future plans:

- Create more meaningful data display of bioassay-scale studies
- Capture and highlight NTP conclusions
- Provide mechanism for CEBS user to start with study conclusion THEN visualize the underlying raw data
- Provide the user with a list of chemicals (and conditions) that produce a particular phenotype
- Provide mechanism to compare and subset lists of chemicals

Thanks!

- DIR scientists and contractors involved in development of CEBS
- Beth Bowden et al. (XML format of legacy NTP data)
- Mike Rowley and Rachel Frawley (NTP microarray data)

- SRA Contractors Asif Rashid and Hui Gong
  - Design and implementation of CEBS and associated tools